

RESEARCH ARTICLE

Open Access

Copy number variants (CNVs) analysis in a deeply phenotyped cohort of individuals with intellectual disability (ID)

Ying Qiao^{1,2†}, Eloi Mercier^{3†}, Jila Dastan², Jane Hurlburt², Barbara McGillivray², Albert E Chudley⁴, Sandra Farrell⁵, Francois P Bernier⁶, ME Suzanne Lewis^{7*}, Paul Pavlidis^{3*} and Evica Rajcan-Separovic^{1*}

Abstract

Background: DNA copy number variants (CNVs) are found in 15% of subjects with ID but their association with phenotypic abnormalities has been predominantly studied in smaller cohorts of subjects with detailed yet non-systematically categorized phenotypes, or larger cohorts (thousands of cases) with smaller number of generalized phenotypes.

Methods: We evaluated the association of *de novo*, familial and common CNVs detected in 78 ID subjects with phenotypic abnormalities classified using the Winter-Baraitser Dysmorphology Database (WBDD) (formerly the London Dysmorphology Database). Terminology for 34 primary (coarse) and 169 secondary (fine) phenotype features were used to categorize the abnormal phenotypes and determine the prevalence of each phenotype in patients grouped by the type of CNV they had.

Results: In our cohort more than 50% of cases had abnormalities in primary categories related to head (cranium, forehead, ears, eye globes, eye associated structures, nose) as well as hands and feet. The median number of primary and secondary abnormalities was 12 and 18 per subject, respectively, indicating that the cohort consisted of subjects with a high number of phenotypic abnormalities (median De Vries score for the cohort was 5). The prevalence of each phenotypic abnormality was comparable in patients with *de novo* or familial CNVs in comparison to those with only common CNVs, although a trend for increased frequency of cranial and forehead abnormalities was noted in subjects with rare *de novo* and familial CNVs. Two clusters of subjects were identified based on the prevalence of each fine phenotypic feature, with an average of 28.3 and 13.5 abnormal phenotypes/subject in the two clusters respectively ($P < 0.05$).

(Continued on next page)

* Correspondence: suzanne.lewis@ubc.ca; paul@chibi.ubc.ca; eseparovic@cw.bc.ca

†Equal contributors

⁷Department of Medical Genetics, BC Child and Family Research Institute and BC Children's and Women's Health Center, University of British Columbia, C234, 4500 Oak Street, Vancouver, BC V6H 3N1, Canada

³Centre for High-throughput Biology, 177 Michael Smith Laboratories, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

¹Department of Pathology (Cytogenetics), BC Child and Family Research Institute, University of British Columbia (UBC), 950 West 28th, Room 3060, Vancouver, BC V5Z 4H4, Canada

Full list of author information is available at the end of the article

(Continued from previous page)

Conclusions: Our study is a rare example of using standardized, deep morphologic phenotype clustering with phenotype/CNV correlation in a cohort of subjects with ID. The composition of the cohort inevitably influences the phenotype/genotype association, and our studies show that the influence of the *de novo* CNVs on the phenotype is less obvious in cohorts consisting of subjects with a high number of phenotypic abnormalities. The outcome of phenotype/genotype analysis also depends on the choice of phenotypes assessed and standardized phenotyping is required to minimize variability.

Keywords: Intellectual disability (ID), Copy number variants (CNVs), Phenotype/genotype analysis, Clustering of phenotypes

Background

Intellectual disability (ID) has an overall prevalence of 1–3% [1,2] and is characterized by considerable genetic and phenotypic heterogeneity. Single gene and chromosomal disorders are considered the cause of ID in 7–37% of cases [3], while submicroscopic gains and losses (DNA copy number variants (CNVs)) occur in a further 5–15% of cases [4,5]. Screening for CNVs using chromosome microarrays is now routinely performed in subjects with ID and databases of CNVs identified in subjects with ID or controls facilitate CNV interpretation (e.g. Database of Chromosomal Imbalance and Phenotype in Human Using Ensemble Resources, DECIPHER, <http://decipher.sanger.ac.uk/>, or Database of Genomic Variants, DGV, <http://projects.tcag.ca/variation>, respectively).

The association of unique CNVs with congenital and neurodevelopmental abnormalities has been documented in reports on individual subjects, small groups of similarly affected subjects (for review see [6]) or large cohorts of patients [7–11]. Large cohort studies including thousands of cases have the benefit of assessing the overall characteristics of CNVs (e.g. size, burden) and their influence on phenotype; however, typically, they lack detailed clinical descriptions, with the phenotype derived from referral forms for array testing, rather than from a detailed chart review. Nevertheless, these studies are informative and show that large CNVs (>400 Kb) harboring more genes (i.e. large CNV burden) are more prevalent in cases with more severe developmental phenotypes associated with multiple congenital anomalies (MCA) [7], including craniofacial dysmorphology and cardiac defects, compared to ID without MCA [7,8].

Thus far, the association of the CNV presence/characteristics with a more detailed and systematic clinical description of a larger number of subjects has been rarely performed. Moreover, the various phenotypes selected for analysis mainly are based on a-priori expectations of phenotypes likely to be affected by chromosomal gain or loss. In a pioneering study, De Vries et al. investigated the association of 21 clinical features in 29 and 110 ID subjects with and without subtelomeric region copy number changes, respectively and introduced a five item

checklist (i.e. de Vries Score) to help select ID patients most likely to have submicroscopic subtelomeric rearrangements (family history of ID, prenatal-onset growth retardation, postnatal growth abnormalities, ≥2 facial dysmorphic features, and congenital anomalies). Using this checklist the authors reported a significant correlation of prenatal onset of growth retardation and a positive family history with subtelomeric abnormalities [12].

In contrast, a recent study of >300 ID cases showed that pathogenic CNVs are significantly correlated with congenital heart anomalies among the 23 clinical features analyzed [13]. Prevalence of microcephaly, short stature and low weight was also higher in cases with pathogenic CNVs, but did not reach statistical significance when compared to cases without pathogenic CNVs. In our previous study of 100 cases with autism spectrum disorder (ASD) and ID [14], in which 10 major phenotypes were evaluated, we reported significant prevalence of microcephaly in cases with pathogenic CNVs and a more severe cognitive deficit in comparison to ASD/ID subjects with normal array results [14].

The most recent study correlating CNV types and phenotypes used Human Phenotype Ontology, HPO based standardized phenotyping in a cohort of >5000 ID patients [15]. However, although 34,433 HPO phenotypic features were evaluated the prevalence of only 9 “lumped” features was assessed and reported in different CNV classes (*de novo*, inherited and no rare CNVs). Significantly increased frequency for 7 out the 9 abnormal features was identified (Multiple congenital anomalies, Dysmorphism, Stature, Convulsions, Head circumference, Brain, Heart, Urogenital) in subjects with *de novo* CNVs. The patients were also assessed using a modified de Vries Score which included intellectual disability, prenatal onset of growth retardation, postnatal growth abnormalities, ≥2 dysmorphic facial features and congenital anomalies. A significant prevalence of subjects with >3 De Vries score in both the *de novo* and familial CNV groups in comparison to no rare CNV group was noted in their cohort which had an overall median De Vries score of 2.

Our study was designed to evaluate the association of different types of CNVs and phenotypes found in 78

patients with ID using Winter-Baraitser Dysmorphology Database (WBDD) (formerly the London Dysmorphology Database) (http://www.lmdatabases.com/about_lmd.html) and is to our knowledge the first study using this database for CNV/phenotype correlation analysis. It is also unique because the information on the prevalence of each individually detailed primary and secondary phenotype in subjects with *de novo*, familial and common CNVs was recorded, compared and reported. The patients were also clustered based on the phenotypes and the prevalence of each phenotypic feature in each cluster was assessed.

Methods

Subjects

A total of 78 subjects with ID were included in the analysis, recruited through a network of collaborating clinical geneticists from centers across Canada. The criteria for recruitment were based on the previously published De Vries score of 3 or higher, which resulted in enrolment of predominantly complex cases with an unknown etiology of ID. Phenotypes were collected from patient charts and confirmed by a clinical geneticist and a genetic counsellor for categorical standardization. This subset of patients was chosen based on: a) the use of array platform of similar resolution for analysis (NimbleGen and Agilent); b) availability of detailed clinical information c) previously normal karyotype and Fragile X screening. As controls we used a previously published cohort of 32 cognitively and phenotypically normal subjects (19 females and 13 males) analyzed using the same array platform [16,17]. The use of the DNA from these patients in our cohort was approved by Clinical Ethics Research Board, University of British Columbia. All subjects gave written informed consent for participation in the study and anonymized data were used for the analysis.

Array comparative genomic hybridization (CGH)

Agilent 105 K oligonucleotide array-CGH analysis was performed according to the protocol provided by the company (version 4.0, June 2006, Agilent Technologies, CA, USA) [18]. Feature Extraction software (version 8.1.1.1, Agilent Technologies) rendered image analysis using the manufacturer's recommended settings (CGH_v4_95) and human genome assembly hg18. The minimum absolute average of log2 ratio was 0.25. Higher-resolution 385 K oligonucleotide genome array CGH was performed by courtesy of NimbleGen. Array log2 ratio $> \pm 0.2$ was used for segmentation (region). For both the Agilent and NimbleGen array platforms, 3 consecutive probes were required for a significant CNV call. CNVs from all chromosomes were included in the analysis.

Type of CNVs

All detected CNVs were grouped into 3 subgroups (*de novo*, familial and common CNVs) based on criterion

described previously [19]. Briefly, CNVs completely overlapping with variants reported in at least two studies in the DGV or in our internal controls consisting of cognitively normal subjects [16,17] were considered common CNVs; CNVs that overlapped partially (<50%) or did not overlap with CNVs reported in the DGV or our internal controls were called unique (rare) CNVs and these included *de novo* and familial CNVs. All unique CNVs were confirmed and their origin (parental or *de novo*) determined by a secondary independent method (FISH or qPCR) on available cell pellet or DNA. Common CNVs from DGV v10 for hg18 have been downloaded at http://projects.tcag.ca/variation/tableview.asp?table=DGV_Content_Summary.txt. The database contained 67694 common CNVs at the time of analysis.

Clinical feature classification

The Winter-Baraitser Dysmorphology Database (WBDD) (formerly the London Dysmorphology Database) (http://www.lmdatabases.com/about_lmd.html) was used to systematically categorize the phenotypes of each patient in our cohort. WBDD consists of 34 major clinical features as the primary category, 162 features in the secondary category and numerous further sub-classifications in the tertiary category. We used the primary and secondary categories of WBDD (named as coarse and fine phenotypes, respectively) to classify the phenotypes of our patients. We also slightly modified WBDD by adding Microcephaly and Macrocephaly as secondary categories within the Cranium-primary category (they are listed in the WBDD tertiary category). We also added the following features as separate items in the secondary category: Family history, abnormal pregnancy history, neonatal abnormality, maternal age at birth and paternal age at birth. This resulted in 169 fine phenotypic features.

For our analysis, clinical features that were present in less than 5% (i.e. in less than 4 individuals) or over 95% (i.e. in more than 74 individuals) were excluded. We eventually included 32 coarse phenotypes (after removing Neurology and Pelvis categories with 78/78 and 2/78 individuals, respectively) in the primary category and 80 fine phenotypes in the secondary category. The complete list of coarse and fine phenotypes is presented in Additional file 1: Table S1. The process of phenotype collection from chart review was extremely time-consuming, and to systematically collect the information, we used RedCap (<https://cric.med.ualberta.ca/neurodevnet/>) [20] for both the phenotype and CNV data storage and extraction. It not only shortened our data processing time, but also minimized any mistakes that might be induced in the process.

Statistical analysis

Computing

All computational analysis was done using software R 2.12 for Windows (The R Project for Statistical computing: <http://www.R-681project.org>) [21]. Fisher's exact test was used in comparisons of equality of proportions. CNV size comparison was performed using the Wilcoxon rank-sum test.

Prevalence of clinical features in subjects with different CNV types

Subjects were classified in groups based on the type of CNV present (*de novo*, familial or common). We computed the fraction of each abnormal phenotypic feature in these groups and tested the significance of the difference in the prevalence of each of the phenotypes between subjects with *de novo* versus common CNVs, and familial versus common CNVs using Fisher's exact test (corrected for multiple tests using the Benjamini and Hochberg procedure) [22].

Clustering

We performed a k-means clustering based on a list of 80 fine clinical features. The optimal value for K (number of clusters) was chosen using the Calinski index [23], which represents the ratio of the variance within the clusters and the variance between the clusters. It is similar to an F (ANOVA) statistic. This was performed

by the cascade KM function from the R package vegan 2.0-7 [24].

Results

Characterization of CNVs in subjects with idiopathic ID

The workflow of our study is shown in Figure 1. Using whole genome oligonucleotide microarrays (Agilent 105 K and NimbleGen 385 K), 527 CNVs were identified in 78 subjects with idiopathic ID (on average 7 CNVs/person). CNVs were classified into three subgroups based on the criteria described in Methods. Twenty-one unique *de novo* CNVs, 27 unique familial CNVs and 479 common CNVs were identified in the ID cohort (Table 1 and Additional file 1: Table S2). *De novo* CNVs ranged in size from 310 Kb to 9.7 Mb (2.5 Mb median) and were significantly larger than common CNVs (0.1 Mb median) ($p = 2.3 \times 10^{-11}$, Wilcoxon's rank-sum test). The proportion of duplications and deletions was similar among the categories except for familial CNVs, for which 70% of cases were duplications ($p = 0.002$, as determined by the rank-sum test compared to pooled *de novo* and common CNVs). The proportion of deletions (and thus also duplications) in the common CNVs is similar to that observed in DGV, 59% vs. 64%, respectively. We also examined the overall gene content of the different classes of CNVs. For the purpose of our analysis, genes within 50 Kb of the estimated CNV breakpoints were included. Significantly more genes were found

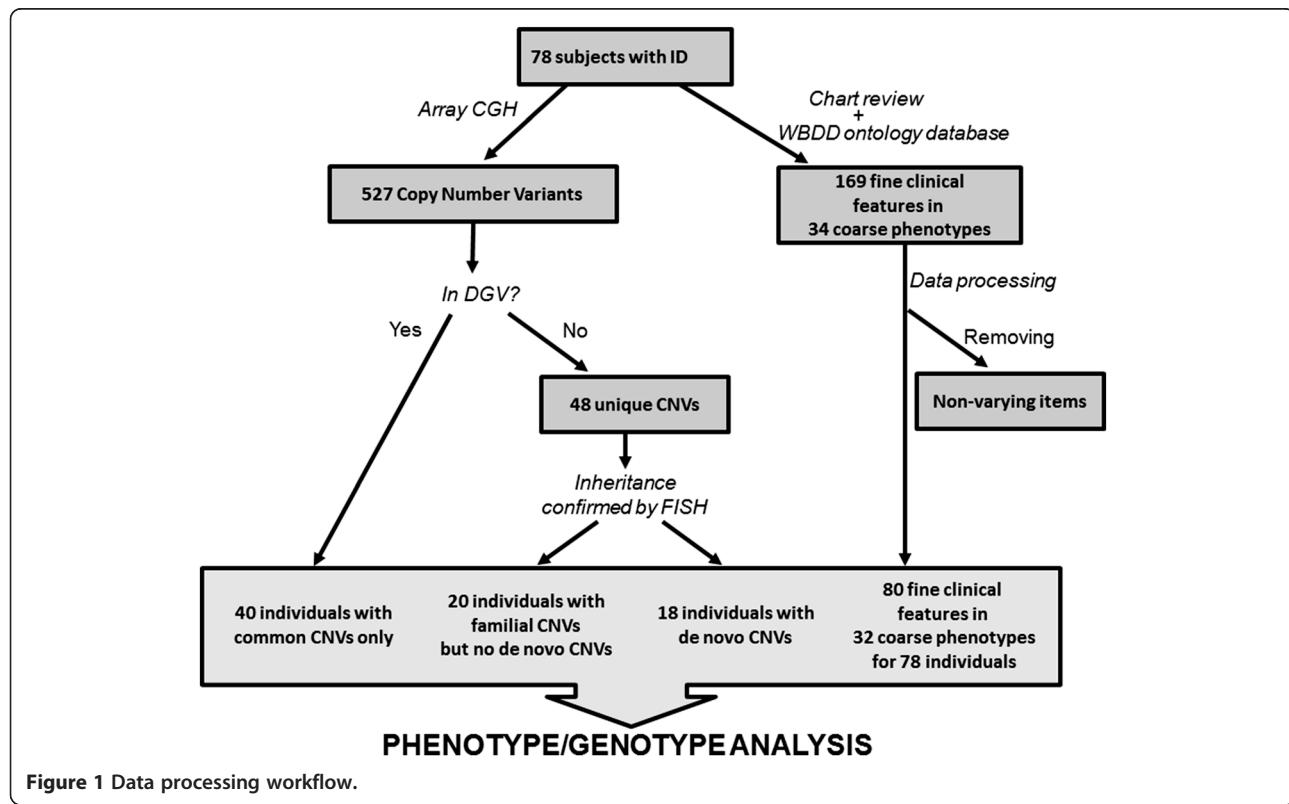


Figure 1 Data processing workflow.

Table 1 CNV features comparison in different CNV types

	Common CNVs	Familial CNVs	De novo CNVs	Overall for the cohort
Number of CNVs	479	27	21	527
Patients	78	22	18	78
CNVs/subject	6.14	1.23	1.17	6.77
Mean CNV size (Mb)	0.42	0.70	3.14	0.54
Median CNVs size (Mb)	0.14	0.43	2.46	0.14
Range of CNV size (Mb)	0.01–9.68	0.03–4.36	0.31–9.78	0.01–9.78
Genes/CNV	5.12	7.04	33.1	6.34
Proportion of large CNVs (>1 Mb)	0.14	0.22	0.71	0.17
Proportion of deletion	0.59	0.30	0.67	0.58
Median modified De Vries Score	5	4.5	5	5

in *de novo* than familial or common CNVs, as would be expected based on the size difference (Table 1).

Six *de novo* CNVs and 5 familial CNVs overlapped with syndromic regions previously described in the DE-CIPHER database (Additional file 1: Table S2). Eighteen of our cases carried *de novo* CNVs (23%), with one case (5%) encompassing two independent *de novo* CNVs (2q23.3 deletion and 10q21.1 deletion) (Additional file 1: Table S2). The slightly higher prevalence of *de novo* CNVs in comparison to the literature could be the effect of enrolment criteria which was based on De Vries scoring system and typically included more phenotypically complex cases. In the unique familial CNV group, 3/22 cases (13%) have 2–3 familial CNVs. There are 2 cases having both a *de novo* and a familial CNV.

Clinical phenotypes classification

Patient records including detailed consult letters were reviewed to categorize the clinical information in 34 coarse and 169 fine clinical features for each subject, using the Winter-Baraitser Dysmorphology Database (WBDD) (www.lmdatabases.com/about_lmd.htm) (Additional file 1: Table S1). The phenotypic categories were slightly modified (see Methods for details) by removing from analysis non-varying phenotypes (e.g. present or absent in more than 95% of the subjects). In addition, we included categories such as prenatal and family history (see Methods for details), and obtained a working set of 80 “fine” phenotypes within 32 “coarse” categories corresponding to the WBDD ontology. The median number of coarse and fine abnormalities was 12 and 18 per subject, respectively.

Other than the neurology class (100%), the most prevalent phenotypes in our cohort, present in >50% of cases, were abnormalities of the head, such as abnormalities of the cranium (72%), ears (68%), eyes (67%) and nose (64%), as well as abnormalities of hands (69%) and feet (65%) (Figure 2, Additional file 1: Table S1). The median De Vries Score (Vulto-van Silfhout et al. [15])

was used to ascertain the severity of phenotypes in our cohort. Seventy-five out of 78 cases (96%) have a score ≥ 3 and the median De Vries score of the whole cohort is 5.

Phenotype/genotype analysis

CNV type/phenotype data for all patients individually are presented in Additional file 1: Table S3. To explore the relationship between the abnormal phenotypes and presence of *de novo*, familial and common CNVs we examined for patients in the 3 CNV groups the median number of coarse and fine abnormalities, the modified de Vries score and the prevalence of each phenotypic feature. We also compared the median de Vries score in subjects with deletions and duplications. Finally, presence of patterns of CNV/phenotype associations for the whole cohort was explored using clustering analysis.

The median number of coarse abnormalities in sub-groups of patients with *de novo*, familial, and common CNVs was 12.5, 10.5, and 14.5 while for fine phenotypes, it was 17.5, 14.5, and 19 for each sub-group, respectively. The modified De Vries score was 5, 4.5 and 5 for sub-groups with *de novo* CNVs, unique familial and common CNVs, respectively. No statistically significant difference was found for the prevalence of any of the phenotypes in different CNV groups after corrections for multiple tests (Fisher's exact test, corrected for multiple tests). However, our data showed that among the phenotypes present in >20% of cases, abnormalities of the forehead and cranium were more prevalent in subjects with *de novo* than common CNVs (Figure 3). When 80 fine phenotypes were considered, abnormalities of forehead (i.e., shape, height, prominence etc.), of brain (structural anomalies), deafness (conductive and sensorineural) and macrocephaly (OFC >98%) were more prevalent in cases with *de novo* than with common CNVs, although this was not significant after multiple test corrections (Additional file 2: Figure S1).

Similarly, a higher prevalence of forehead anomalies was noted in subjects with familial CNVs when coarse

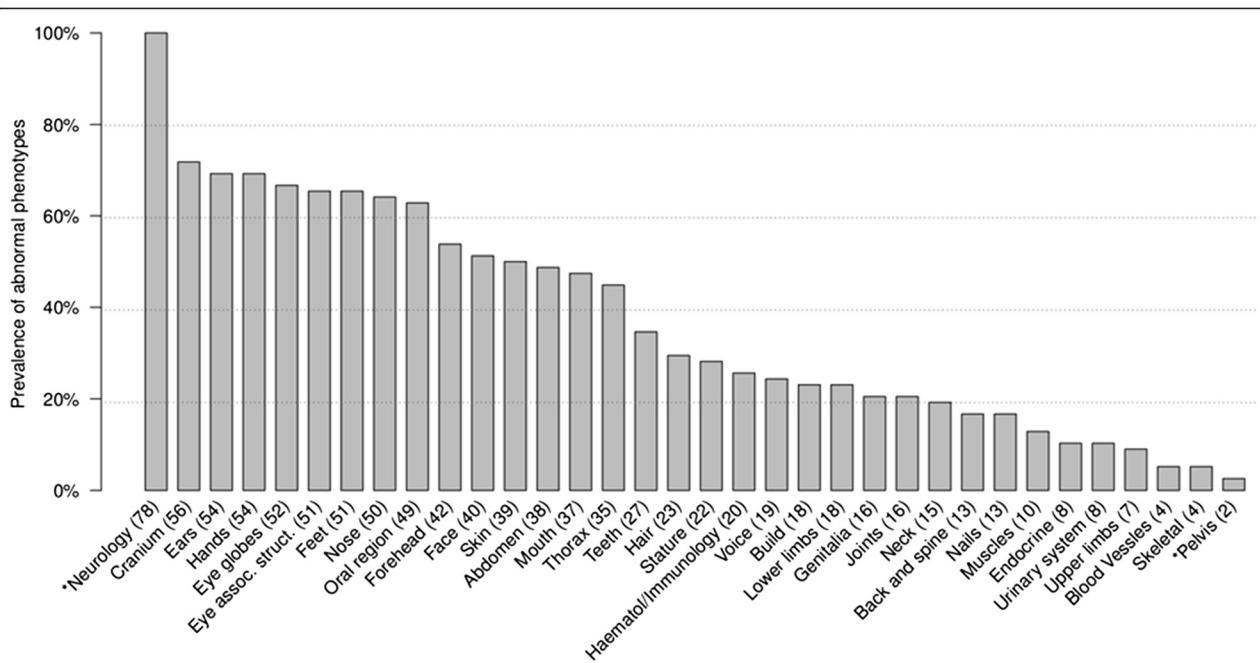


Figure 2 Prevalence of abnormal coarse phenotypes. Thirty-four coarse phenotypes were evaluated among our 78 patients based on WBDD criteria (see Additional file 1: Table S1 for the whole term of each phenotype). *indicates phenotype with >95% or <5% prevalence in the cohort which was removed in the statistical analysis.

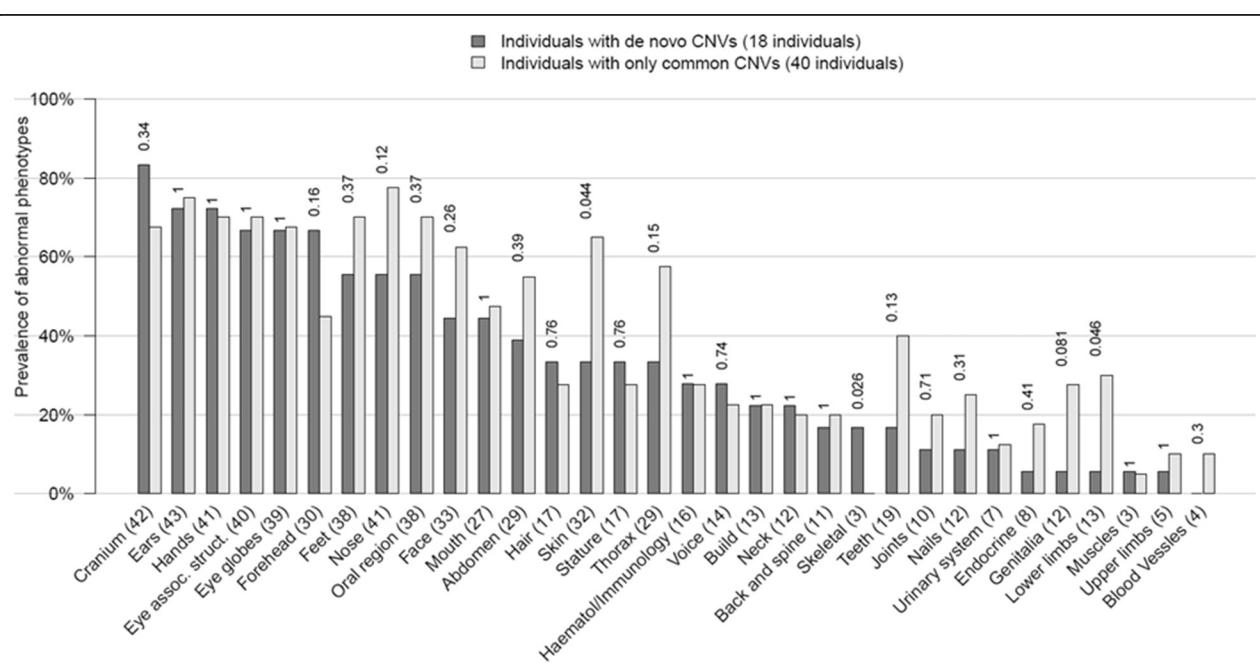


Figure 3 Phenotype and de novo CNV association analysis. Prevalence of the abnormality of each of the coarse phenotypes in individuals with de novo CNVs (18 cases) compared to individuals with only common CNVs (40 cases). The phenotypes with a prevalence >95% or <5% in the whole cohort (78 cases) were excluded from calculation.

phenotypes were analyzed (Figure 4). When 80 fine phenotypes were considered, number of cases with family history of ID, and with forehead anomalies was higher in the familial than common CNV group, and muscle abnormalities were seen in ~5 times more cases with familial CNVs. However, these frequencies did not reach significant levels after corrections for multiple tests (Additional file 3: Figure S2). The type of CNV (deletion or duplication) slightly affected the severity of the phenotype based on the modified De Vries score (score of 5.5 for deletions and 4.6 for duplications).

Finally, to explore the association of clinical phenotypes with CNV subtypes more generally, K-means clustering analysis was performed on patients based on the 80 fine phenotypes. The optimal number of clusters was computationally determined to be two (see Methods). Individuals belonging to the first cluster had significantly more phenotypic abnormalities (mean 28.3/subject) than those from the second cluster (mean 13.5/subject; $p = 2.7 \times 10^{-12}$; Wilcoxon rank-sum test) (Figure 5). 24 out of 80 phenotypes were significantly more prevalent in cluster 1 compared to cluster 2 ($P < 0.05$, Fisher's exact test after multiple test correction) (Figure 5 and Additional file 1: Table S4). We stress that differences in phenotypes between the clusters are expected since the clustering is based on the phenotypes. However, neither the number of total CNVs, the number of de novo or familial CNVs, nor CNV size segregated with the clusters.

Discussion

This report contributes a unique exploration of the association of detailed phenotypic categories applied from the LDD with *de novo*, familial and common CNV subtypes, to systematically record, compare and report primary and secondary phenotypic abnormalities in 78 ID subjects. Our cohort consisted of subjects with a high number of phenotypic abnormalities with a median of 12 for primary and 18 for secondary features/subject. This was also reflected in a high median modified De Vries score of 5 for the whole cohort. We did not detect significant prevalence for any of the phenotypes in subjects with unique *de novo* or familial CNVs in comparison to those with common CNVs only, and it is possible that the high and comparable severity of the phenotype in three CNV subgroups in our cohort eliminated the CNV impact. Nevertheless, we noted higher prevalence of several abnormalities in the unique (*de novo* and familial) CNV subgroup in comparison to the common CNV subgroup (e.g. forehead abnormalities) while in subjects with only common CNVs, abnormalities of skin and thorax were present almost 2 times more frequently than in subjects with *de novo* or familial CNVs.

There are very few previous studies that correlate 10 ~ 23 phenotypic features in subjects with ID with the presence or absence of submicroscopic genomic changes. No consistent results were found among these studies regarding the specific phenotypes significantly prevalent in each

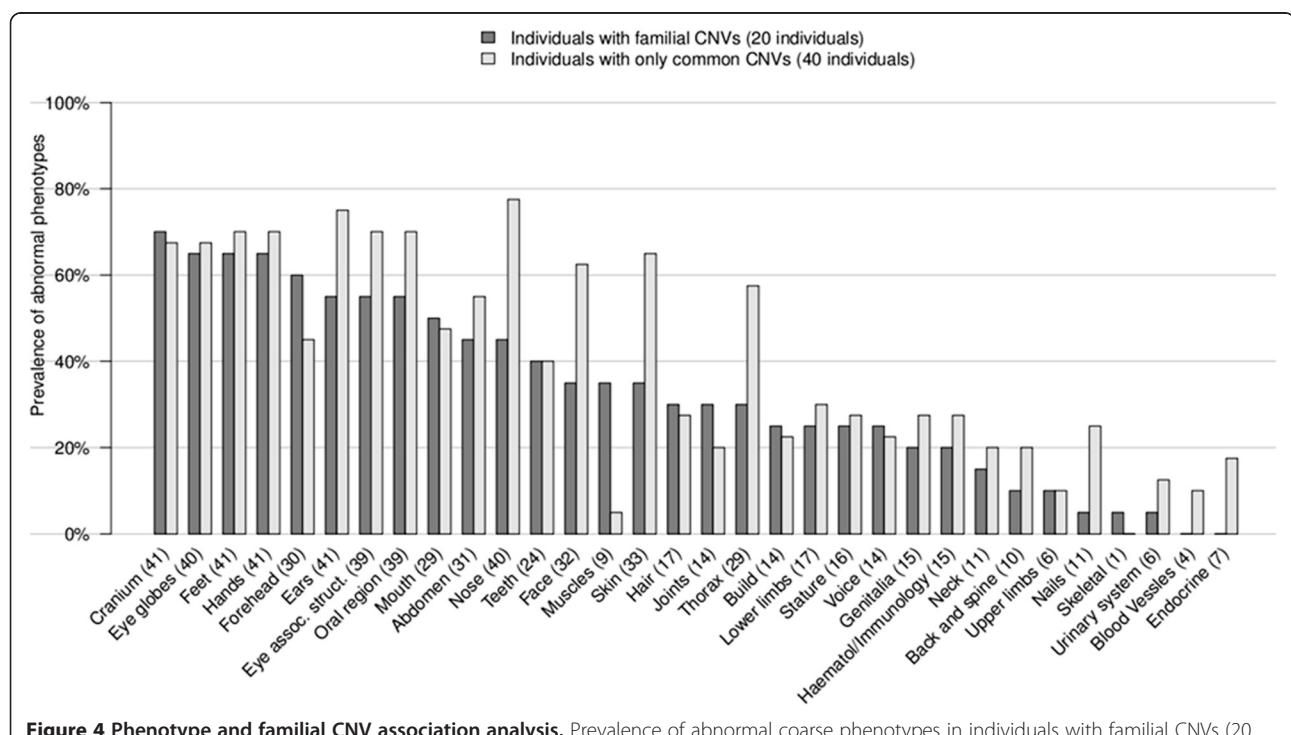
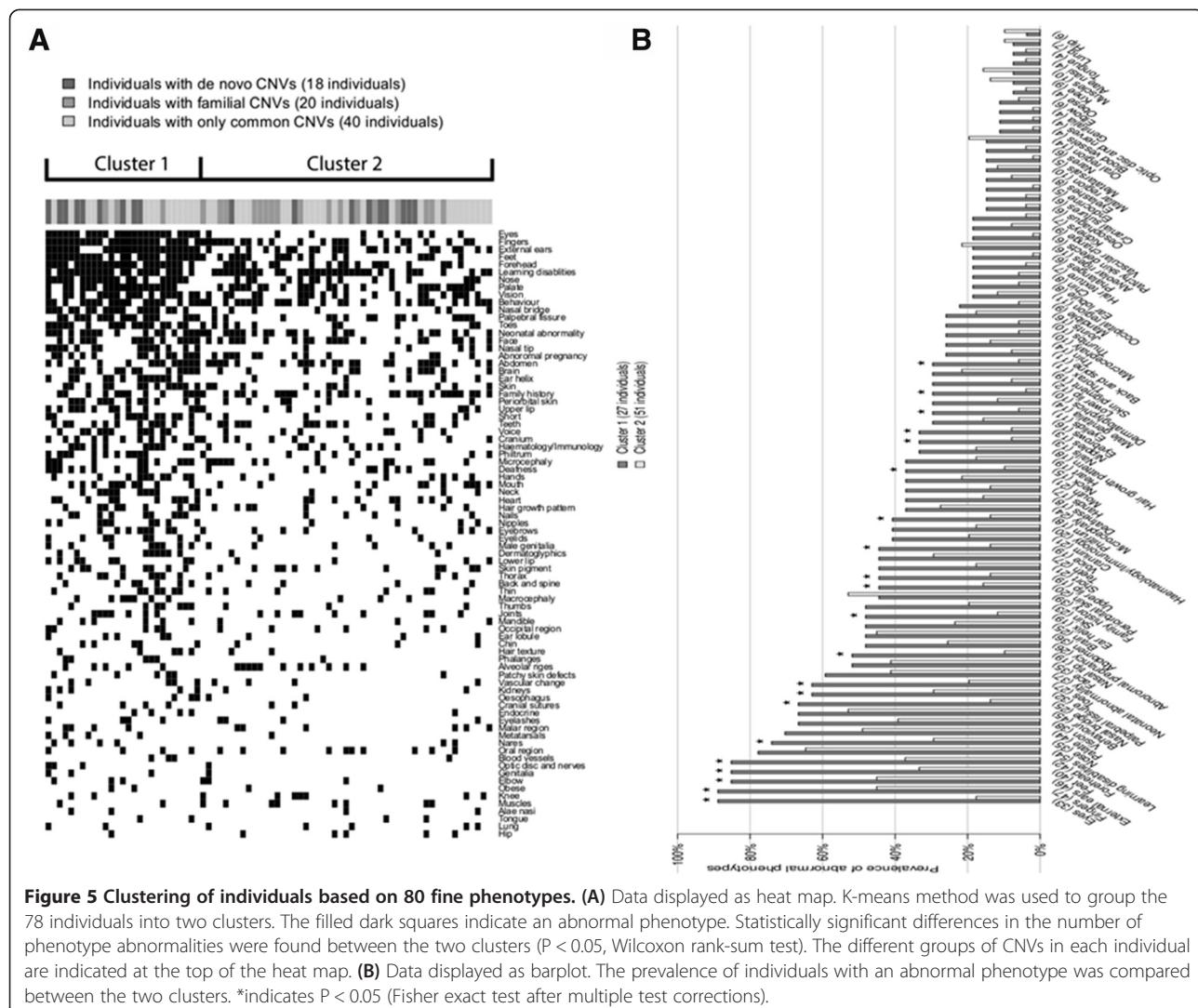


Figure 4 Phenotype and familial CNV association analysis. Prevalence of abnormal coarse phenotypes in individuals with familial CNVs (20 cases) compared with those containing only common CNVs (40 cases). Two individuals with both *de novo* and familial CNVs were removed from the analysis. The phenotypes with a prevalence >95% or <5% in the whole cohort (78 cases) were excluded from calculation.



cohort. De Vries et al. reported a significantly higher incidence of prenatal abnormalities and positive family history of ID in children with subtelomeric abnormalities than in patients without subtelomeric defects [12], while our previous study of ASD/ID subjects [14] noted that microcephaly and severity of ID were more significantly present in cases with pathogenic CNVs in comparison to cases without pathogenic CNVs. More recently, significantly higher prevalence of heart abnormalities in ID subjects with clinically relevant CNVs or chromosome abnormalities, was noted by Shoukier et al. [13], while statistical difference in the prevalence of microcephaly and short stature was not reported between the groups. Of note, higher prevalence of macrocephaly, epilepsy and short stature was reported in subjects with pathogenic CNVs. The most recent study by Vulto-van Silfhout et al. identified facial dysmorphisms, abnormal head circumference, central nervous system anomalies, heart anomalies, urogenital anomalies and modified De Vries scores ≥ 3 to occur at significantly

higher frequency in subjects with de novo CNVs based on assessment of >5000 subjects phenotyped using HPO.

Possible reasons for discrepancy between studies include selection biases in ID subjects that had array testing (study cohorts). For example our cohort had a median De Vries score of 5, while for the cohort of Vulto-van Silfhout et al. the median score was 2. In addition, differences in the classification of CNVs exist between studies; for example Shoukier et al. included as pathogenic CNVs large scale chromosome abnormalities and syndromic and familial CNVs, while Vulto-van Silfhout excluded syndromic CNVs caused by LCRs and divided the patients based on presence of rare de novo, familial or no rare CNVs. Finally, differences in available/recoded phenotypic characteristics of patient cohorts, differences in the selection of clinical features being evaluated, or the discrepancy in the stringency or type of statistical methods used for data analysis, could be the cause of variable genotype/phenotype associations. In our study, the clinical information was obtained

retrospectively and depended on the classification and description preferences of each of the participating clinical geneticists, and these also could influence the findings. Ideally, the use of a relevant and standardized ontology classification of phenotypes derived from deep phenotyping initiatives will improve phenotype/genotype analyses relevant to scientific discovery and personalized patient management of genomic causes of ID.

The WBDD database catalogues phenotypes systematically by annotation of anatomic regions and systems for the human body. Only the primary and secondary phenotype categories with more concise descriptors were used in our study, to avoid the overwhelming detail of tertiary category designations (mostly absent from patient records). The WBDD is user-friendly and easy to master, with the definition for most of the symptoms provided by the database. However, in our consideration of specific characteristics of patients with ID phenotypes, we found the database presented some limitations. For example, it does not include prenatal information, family history, severity of ID (by IQ or adaptive/functional measures), all of which could offer essential elements of the phenotype in the context of ID. Similarly, some phenotypes commonly described in practice, such as motor delays (oral, fine, gross motor), cranio-facial dysmorphism, microcephaly and macrocephaly, are not listed as isolated items in primary or secondary categories. In addition, the best match for ID is neurology in the primary category, which contains three secondary features: behaviour, learning disabilities and neuro-abnormalities. The WBDD also contains an extended number of features that are rarely reported in ID within the primary categories such as pelvis, voice and skeletal system. A directly targeted, separate, and systematic ontology system for accurate and comprehensive ID phenotypic designations would be beneficial for achieving more accurate phenotype/genotype correlation and clinical translation. This system should have a detailed description of neurodevelopmental features, considering the prevalence of cranial abnormalities in our cohort.

CNVs are only one of the possible sources of genomic variation that can be pathogenic in ID [4,5]. With the advent of whole exome or genome sequencing techniques, novel sequence mutations have been found to play important role in the pathogenesis of ID in cases with or without detected pathogenic CNVs [25-28]. Our clustering analysis allowed us to group subjects in two clusters based on frequencies of abnormalities (median 28 or 13 per subject) and it will be interest to explore the mutation types and frequencies in these two groups of patients in the future. Establishing the functional consequences of gene copy number or sequence changes is also important for the assessment of their impact on the phenotype and studies addressing closer functional and

phenomic linkages are becoming more common [29-33]. Efforts to use a more standardized and detailed phenotyping system in combination with array-CGH, sequencing and gene functional analysis is needed to improve our understanding of phenotype/genotype correlations and optimize their translation into accurate genetic counselling.

Conclusions

Our study uniquely explores the association of *de novo*, familial and common CNV subtypes with detailed phenotypes categorized by a commonly used human phenotype ontology database. Our cohort consisted of cases with a high median number of phenotypic abnormalities in all CNV subgroups which possibly resulted in no significant difference in the frequency of any of the studied phenotypes between the CNV sub-groups. Nevertheless, our study provides a detailed comprehensive and systematic cross-section of the frequencies of primary and secondary phenotypes in CNV sub-groups based on WBDD. We found WBDD to be user-friendly and easy to master with the definition for most of the symptoms provided by the database. Wider use of standardized and detailed phenotyping systems in combination with current whole genome analyses, including chromosome arrays and whole genome sequencing, is needed for achieving more accurate phenotype/genotype correlation and clinical translation.

Additional files

Additional file 1: Table S1. WBDD phenotype frequency in our cohort. Table S2 De novo and familial CNVs detected in the cohort (hg18). Table S3 Phenotype data in 78 cases with ID. Table S4. Prevalence of abnormal fine phenotypes in two clusters.

Additional file 2: Figure S1. Prevalence of secondary phenotypes in *de novo* CNV group. Prevalence of abnormal fine phenotypes in individuals with *de novo* CNVs (18 cases) compared with those containing only common CNVs (40 cases). The phenotypes with prevalence >95% or <5% in the whole cohort (78 cases) were excluded from calculation.

Additional file 3: Figure S2. Prevalence of secondary phenotypes in familial CNV group. Prevalence of abnormal fine phenotypes in individuals with familial CNVs (20 cases) compared with those containing only common CNVs (40 cases). Two individuals with both *de novo* and familial CNVs were removed from the analysis. The phenotypes with a prevalence >95% or <5% in the whole cohort (78 cases) were excluded from calculation.

Competing interest

We declare no conflict of interest in our manuscript titled as "Copy Number Variant analysis in a deeply phenotyped cohort of individuals with Intellectual Disability (ID)".

Authors' contributions

YQ performed genetic and clinical data acquisition and analysis, and drafted the manuscript; EM performed statistical and bioinformatics analyses, and drafted the manuscript; JD participated in phenotype data analysis; BM, AC, SF and FB recruited clinical cases and reviewed the manuscript; SML recruited clinical cases, supervised phenotype data analysis and reviewed the manuscript; PP supervised statistical and bioinformatics analyses, and reviewed the manuscript; ERS supervised and designed the study, helped

with data interpretation, and critically revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by funding from the Canadian Institutes for Health Research (CIHR) (MOP 74502; PI: ERS). PP was supported by a career award from the Michael Smith Foundation for Health Research, a CIHR New Investigator award, the Canadian Foundation for Innovation, and the National Institutes of Health (GM076990). MESL and ERS are Career Scholars supported by the Michael Smith Foundation for Health Research. We thank Elodie Portales-Casamar and the NeuroDevNet Neuroinformatics Core for assistance with Redcap.

Author details

¹Department of Pathology (Cytogenetics), BC Child and Family Research Institute, University of British Columbia (UBC), 950 West 28th, Room 3060, Vancouver, BC V5Z 4H4, Canada. ²Department of Medical Genetics, BC Child and Family Research Institute, UBC, Vancouver, BC V6H 3N1, Canada. ³Centre for High-throughput Biology, 177 Michael Smith Laboratories, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. ⁴Department of Human Genetics, University of Manitoba Children's Hospital, Winnipeg, Manitoba R3A 1R9, Canada. ⁵The Trillium Health Partners, 2200 Eglinton Avenue West, Mississauga, ON L5M 2N1, Canada. ⁶Department of Medical Genetics, University of Calgary, Calgary, Alberta, Canada. ⁷Department of Medical Genetics, BC Child and Family Research Institute and BC Children's and Women's Health Center, University of British Columbia, C234, 4500 Oak Street, Vancouver, BC V6H 3N1, Canada.

Received: 5 February 2014 Accepted: 3 July 2014

Published: 16 July 2014

References

- Chelly J, Khelfaoui M, Francis F, Cherif B, Bienvenu T: **Genetics and pathophysiology of mental retardation.** *Eur J Hum Genet* 2006, **14**:701–713.
- Roeleveld N, Zielhuis GA, Gabreels F: **The prevalence of mental retardation: a critical review of recent literature.** *Dev Med Child Neurol* 1997, **39**:125–132.
- Curry CJ, Stevenson RE, Aughton D, Byrne J, Carey JC, Cassidy S, Cunniff C, Graham JM Jr, Jones MC, Kaback MM, Moeschler J, Schaefer GB, Schwartz S, Tarleton J, Opitz J: **Evaluation of mental retardation: recommendations of a Consensus Conference: American College of Medical Genetics.** *Am J Med Genet* 1997, **72**:468–477.
- Koolen DA, Pfundt R, de Leeuw N, Hehir-Kwa JY, Nillesen WM, Neefs I, Scheltinga I, Sisternmans E, Smeets D, Brunner HG, van Kessel AG, Veltman JA, de Vries BB: **Genomic microarrays in mental retardation: a practical workflow for diagnostic applications.** *Hum Mutat* 2009, **30**:283–292.
- Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, Church DM, Crolla JA, Eichler EE, Epstein CJ, Fawcett WA, Feuk L, Friedman JM, Hamosh A, Jackson L, Kaminsky EB, Kok K, Krantz ID, Kuhn RM, Lee C, Ostell JM, Rosenberg C, Scherer SW, Spinner NB, Stavropoulos DJ, Tepperberg JH, Thorland EC, Vermeesch JR, Waggoner DJ, Watson MS, et al: **Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies.** *Am J Hum Genet* 2010, **86**:749–764.
- Sahoo GS, Butterworth AS, Sanderson S, Shaw-Smith C, Higgins JP, Burton H: **Array CGH in patients with learning disability (mental retardation) and congenital anomalies: updated systematic review and meta-analysis of 19 studies and 13,926 subjects.** *Genet Med* 2009, **11**:139–146.
- Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, Abdel-Hamid H, Bader P, McCracken E, Niayoz D, Leppig K, Thiese H, Hummel M, Alexander N, Gorski J, Kussmann J, Shashi V, Johnson K, Rehder C, Ballif BC, Shaffer LG, Eichler EE: **A copy number variation morbidity map of developmental delay.** *Nat Genet* 2011, **43**:838–846.
- Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, Vu TH, Shaffer N, Bernier R, Ferrero GB, Silengo M, Warren ST, Moreno CS, Fichera M, Romano C, Raskind WH, Eichler EE: **Relative burden of large CNVs on a range of neurodevelopmental phenotypes.** *PLoS Genet* 2011, **7**:e1002334.
- Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, Moreno-De-Luca D, Moreno-De-Luca A, Mulle JG, Warren ST, Richard G, Compton JG, Fuller AE, Gliem TJ, Huang S, Collinson MN, Beal SJ, Ackley T, Pickering DL, Golden DM, Aston E, Whitby H, Shetty S, Rossi MR, Rudd MK, South ST, Brothman AR, Sanger WG, Iyer RK, Crolla JA, et al: **An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities.** *Genet Med* 2011, **13**:777–784.
- Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A, Filipink RA, McConnell JS, Angle B, Meschino WS, Nezariati MM, Asamoa A, Jackson KE, Gowans GC, Martin JA, Carmany EP, Stockton DW, Schnur RE, Penney LS, Martin DM, Raskin S, Leppig K, Thiese H, Smith R, Aberg E, Niayoz D, Escobar LF, El-Khechen D, Johnson KD, Lebel RR, et al: **Phenotypic heterogeneity of genomic disorders and rare copy-number variants.** *N Engl J Med* 2012, **367**:1321–1331.
- Coe BP, Girirajan S, Eichler EE: **A genetic model for neurodevelopmental disease.** *Curr Opin Neurobiol* 2012, **22**:829–836.
- de Vries BB, White SM, Knight SJ, Regan R, Homfray T, Young ID, Super M, McKeown C, Splitter M, Quarrell OW, Trainer AH, Niermeijer MF, Malcolm S, Flint J, Hurst JA, Winter RM: **Clinical studies on submicroscopic subtelomeric rearrangements: a checklist.** *J Med Genet* 2001, **38**:145–150.
- Shoukier M, Klein N, Auber B, Wickert J, Schroder J, Zoll B, Burfeind P, Bartels I, Alsat EA, Lingen M, Grzmil P, Schulze S, Keyser J, Weise D, Borchers M, Hobbebrunken E, Robl M, Gartner J, Brockmann K, Zirn B: **Array CGH in patients with developmental delay or intellectual disability: are there phenotypic clues to pathogenic copy number variants?** *Clin Genet* 2012, **83**:53–65.
- Qiao Y, Riendeau N, Koochek M, Liu X, Harvard C, Hildebrand MJ, Holden JJ, Rajcan-Sepovic E, Lewis ME: **Phenomic determinants of genomic variation in autism spectrum disorders.** *J Med Genet* 2009, **46**:680–688.
- Vulto-van Silfhout AT, Hehir-Kwa JY, van Bon BW, Schuurs-Hoeijmakers JH, Meader S, Hellebrekers CJ, Thoonen JJ, de Brouwer AP, Brunner HG, Webber C, Pfundt R, de Leeuw N, de Vries BB: **Clinical significance of de novo and inherited copy-number variation.** *Hum Mutat* 2013, **34**:1679–1687.
- Rajcan-Sepovic E, Diego-Alvarez D, Robinson WP, Tyson C, Qiao Y, Harvard C, Fawcett C, Kalousek D, Philipp T, Somerville MJ, Stephenson MD: **Identification of copy number variants in miscarriages from couples with idiopathic recurrent pregnancy loss.** *Hum Reprod* 2010, **25**:2913–2922.
- Rajcan-Sepovic E, Qiao Y, Tyson C, Harvard C, Fawcett C, Kalousek D, Stephenson M, Philipp T: **Genomic changes detected by array CGH in human embryos with developmental defects.** *Mol Hum Reprod* 2010, **16**:125–134.
- Fan YS, Jayakar P, Zhu H, Barbouth D, Sacharow S, Morales A, Carver V, Benke P, Mundy P, Elsa L: **Detection of pathogenic gene copy number variations in patients with mental retardation by genome-wide oligonucleotide array comparative genomic hybridization.** *Hum Mutat* 2007, **28**:1124–1132.
- Qiao Y, Tyson C, Hrychak M, Lopez-Rangel E, Hildebrand J, Martell S, Fawcett C, Kasmara L, Calli K, Harvard C, Liu X, Holden JJ, Lewis SM, Rajcan-Sepovic E: **Clinical application of 2.7 M Cytogenetics array for CNV detection in subjects with idiopathic autism and/or intellectual disability.** *Clin Genet* 2013, **83**:145–154.
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG: **Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support.** *J Biomed Inform* 2009, **42**:377–381.
- R Core Team: **A Language and Environment for Statistical Computing.** R Foundation for Statistical Computing 2013, Vienna, Austria: <http://www.R-project.org/>.
- Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Stat Soc Ser B* 1995, **57**:289–300.
- Calinski T, Harabasz J: **A dendrite method for cluster analysis.** *Commun Statist* 1974, **3**:1–27.
- Oksanen J, Blanchet G, Kindt R, Legendre P, Minchin P, O'Hara RB, Simpson GL, Solymos P, Stevens H, Wagner H: **R package version 2.0-7: vegan: Community Ecology Package** 2013, <http://CRAN.R-project.org/package=vegan>.
- Classen CF, Riehmer V, Landwehr C, Kosfeld A, Heilmann S, Scholz C, Kabisch S, Engels H, Tierling S, Zivcicnak M, Schacherer F, Haffner D, Weber RG: **Dissecting the genotype in syndromic intellectual disability using whole exome sequencing in addition to genome-wide copy number analysis.** *Hum Genet* 2013, **132**:825–841.
- de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, Vulto-van Silfhout AT, Koolen DA, de Vries P, Gilissen C, del Rosario M,

Hoischen A, Scheffer H, de Vries BB, Brunner HG, Veltman JA, Vissers LE: **Diagnostic exome sequencing in persons with severe intellectual disability.** *N Engl J Med* 2012, **367**:1921–1929.

27. Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, Schwarzmayr T, Albrecht B, Bartholdi D, Beygo J, Di Donato N, Dufke A, Cremer K, Hempel M, Horn D, Hoyer J, Josten P, Röpke A, Moog U, Riess A, Thiel CT, Tzschach A, Wiesener A, Wohlleber E, Zweier C, Ekici AB, Zink AM, Rump A, Meisinger C, Grallert H, Sticht H, et al: **Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study.** *Lancet* 2012, **380**:1674–1682.

28. Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M, van Bon BW, Hoischen A, de Vries BB, Brunner HG, Veltman JA: **A de novo paradigm for mental retardation.** *Nat Genet* 2010, **42**:1109–1112.

29. Luo R, Sanders SJ, Tian Y, Voineagu I, Huang N, Chu SH, Klei L, Cai C, Ou J, Lowe JK, Hurles ME, Devlin B, State MW, Geschwind DH: **Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders.** *Am J Hum Genet* 2012, **91**:38–55.

30. Colnaghi R, Carpenter G, Volker M, O'Driscoll M: **The consequences of structural genomic alterations in humans: genomic disorders, genomic instability and cancer.** *Semin Cell Dev Biol* 2011, **22**:875–885.

31. Outwin E, Carpenter G, Bi W, Withers MA, Lupski JR, O'Driscoll M: **Increased RPA1 gene dosage affects genomic stability potentially contributing to 17p13.3 duplication syndrome.** *PLoS Genet* 2011, **7**:e1002247.

32. Kerzendorfer C, Hannes F, Colnaghi R, Abramowicz I, Carpenter G, Vermeesch JR, O'Driscoll M: **Characterizing the functional consequences of haploinsufficiency of NELF-A (WHSC2) and SLBP identifies novel cellular phenotypes in Wolf-Hirschhorn syndrome.** *Hum Mol Genet* 2012, **21**:2181–2193.

33. Harvard C, Strong E, Mercier E, Colnaghi R, Alcantara D, Chow E, Martell S, Tyson C, Hrychak M, McGillivray B, Hamilton S, Marles S, Mhanni A, Dawson AJ, Pavlidis P, Qiao Y, Holden JJ, Lewis SM, O'Driscoll M, Rajcan-Separovic E: **Understanding the impact of 1q21.1 copy number variant.** *Orphanet J Rare Dis* 2011, **6**:54–66.

doi:10.1186/1471-2350-15-82

Cite this article as: Qiao et al.: Copy number variants (CNVs) analysis in a deeply phenotyped cohort of individuals with intellectual disability (ID). *BMC Medical Genetics* 2014 15:82.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

